Proviso with pseudowords: Summary of five pilot studies

1. Introduction

The proviso problem is a problem of *non-uniformity* in presupposition projection: presuppositions of elementary clauses are sometimes inherited wholesale when the clause is embedded under a truth-conditional operator (1), but sometimes, the presuppositions associated with complex seem to be weaker than those of its simplex constituents (2a).

(1) Sam didn't bring his guitar.
—> Sam has a guitar.

(2) a. If Sam is a musician, he will bring his guitar to the party.
—> If Sam is a musician, he has a guitar.
b. If Sam is tired, he will bring his guitar to the party.
—> Sam has a guitar.

In particular, the felt presupposition associated with (2a) seems to be a *conditional* one, whereas the one associated with (2b), whose sentence takes a similar form than (2a), is non-conditional and stronger. The conditional inference in (2a) is predicted by various theories of projection, which either encode this projection behavior (i) directly as a lexical property of connectives (e.g. Heim 1983), (ii) a consequence of the way presupposition failure is repaired within trivalent system (George 2008, Fox 2008), or (iii) a consequence of how contexts evolve over the course of evaluating a complex sentence (Karttunen 1974, Soames 1982, Schlenker 2009, 2011). Within dynamic accounts, a standard response to the strengthened, non-conditional inference that we obtain in (2b) is to supplement the existing system of presupposition projection with a strengthening mechanism, the details of which vary across different accounts. Recent works continue to address this problem (van Rooij 2007; Singh 2007; Schlenker 2011; Lassiter 2012; Mayr & Romoli 2016; Mandelkern 2016; Mandelkern & Rothschild 2018; Francez 2019; Winter 2019).

As Singh (2007, 2008) and Schlenker (2011) point out, the proviso problem can be decomposed into two subproblems that can be tackled separately. More specifically, any complete theory of presupposition projection should be equipped with mechanisms to generate both conditional and non-conditional inferences, and also it should be able to predict *which* inference obtains *when*. These are known as the Strengthening problem and the Selection problem, respectively:

(3) **The Strengthening Problem**: How are the different inferences generated?
The Selection Problem: How is the inference that actually obtains selected?

While the details of how the strengthening mechanism is implemented vary across different accounts, there is a widely shared intuition that the (un)availability of a strengthened reading is regulated by some notion of **(in)dependence** or **(ir)relevance**:

specifically, what distinguishes (2a) from (2b) is that in (2a), *p* and *q̲* are dependent, since our world knowledge tells us that someone who is a musician is likely to have a guitar; in (2b), *p* and *q* are not obviously dependent or related to each other, since someone being tired generally has no bearing on whether or not they should have a guitar. Generalizing from this contrast in (2a-b), for any conditional statement that takes the form *if p then q̲q*, with *q̲* being the presupposition in the consequent clause, the strengthening mechanism kicks in and delivers a non-conditional inference if *p* and *q̲* are independent.[1]

While (in)dependence is often recruited in developing the strengthening mechanism, its notion is not always explicitly defined. Two notions of (in)dependence that have been fleshed out in recent work are van Rooij's (2007) *qualitative* definition, which is based on entailment by input context, and Lassiter's (2012) (and to a lesser extent maybe Schlenker's (2011) too) *probabilistic* definition. It is **not** our goal to tease apart different notions of independence, and as Mandelkern & Rothschild (2018) points out, there is a sense in which qualitative independence is the corollary of probabilistic independence in a qualitative framework. The motivation of our study, to put briefly, is that while we know that *some* notion of independence plays a role behaviorally— based on our intuitions, and an existing study which we briefly discuss below — what underlies this behavioral signature remains much less understood. Is it part of the grammar, which follows from the logic of presupposition projection in conditionals given our semantic theory, or is it primarily post-semantic reasoning based on world knowledge about other people's information states and commitments?

Our goal then, is precisely to divorces issues of world knowledge and plausibility considerations from the interpretation of conditionals, which we hope will give us a real shot at understanding what is the semantic presupposition of these sentences. By doing so, we can focus on the presupposition projection aspect of the problem. Looking ahead, the questions we may be interested in include:

- What is the semantic presupposition projected out of conditionals that take the form of *if p then q̲q*, if we take away world knowledge considerations from these sentences?
  - If we introduce independence relations using pseudowords, do we still obtain strengthened inferences? Or does non-uniformity disappear?
  - That is, do linguistic structures suffice to guarantee strengthening?
- Mayr & Romoli (2016) analyzes the proviso problem as a matter of truth conditional ambiguity. In particular, they treat the strengthened inference as the result of a bi-conditional derived from *exh* applying to the plain conditional. This provides testable predictions within the current agenda:
  - Do explicit bi-conditonals give rise to uniform non-conditional inferences?

[1] Note that this is not entirely accurate, given what Lassiter (2012) noted about cases of deprobabilization, i.e. pr(q̲|p) < pr(q̲), where the felt presupposition is unconditional even though the crucial independence assumption is clearly not appropriate. However, this suffices for our present purpose, so I set the deprobabilization cases aside for now.

- If we place plain conditionals in an environment that blocks *exh*, do we uniformly get conditional inferences? ("*It is not the case that if p then q*")
- How about different projection environments?
  - Conditionals vs. disjunction
  - Factives, which pose long-standing problems for many accounts

2. Existing work [skippable]

To our knowledge, there has been one published work that attempts to systematically investigates the role of (in)dependence in the proviso problem. Romoli, Sudo, & Snedeker (2012) tested the hypothesis that the conditional inference *p —> q* is more likely to arise when the presupposition *q* in the consequent is *dependent* on the antecedent *p*. The experimental materials in Romoli et al include two critical conditions, **Dependent** and **Independent**, which was determined by results from a separate norming study where participants were given a statement about a monster and were asked whether that statement made it more or less likely that the monster possessed a particular property (*e.g. Googlemorph is flying. Does that make it more likely that he has wings?*). The eight items with the highest scores are used to construct the critical items of the **Dependent** condition, and the eight items whose scores were the *closest to the middle point 2.5* were used to construct the critical items in the **Independent** condition. Notice that this decision already allowed for the possibility that in the **Independent** condition, there was still room for speculating a dependence relation, albeit perhaps a somewhat weak one.

Using a picture selection task (Exp 1), Romoli et al asked participants to read sentences and pick a picture that best matches what the sentences say. The target and control trials involve conditional sentences with a possessive noun phrase. They are followed by either a confirmation of the antecedent (control trials) or a denial of the antecedent (critical trials):

(4) **Dependent**
If Googlemorph is flying, then his wings are big and strong.
a. And Googlemorph is flying. (Control)
b. But Googlemorph is not flying. (Critical)

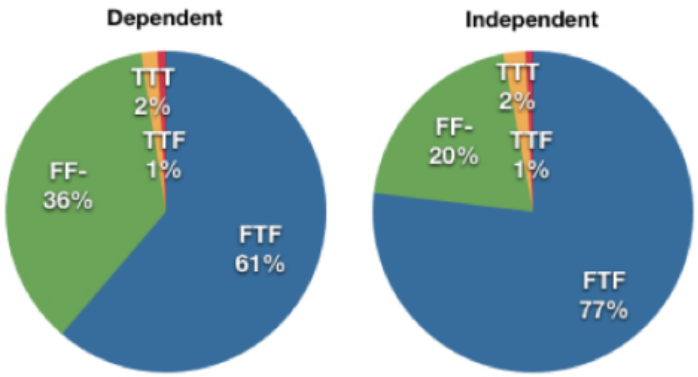
(5) **Independent**
If Googlemorph is drinking orange juice, then his wings are big and strong.
a. And Googlemorph is drinking orange juice. (Control)
b. But Googlemorph is not drinking orange juice. (Critical)

Each item has four pictures, which depict the following four monsters: (A) a monster (TTT) that satisfies all of the antecedent, **the non-conditional presupposition** and the consequent; (B) a monster (TTF) that satisfies the antecedent and **the non-conditional presupposition** but not the consequent; (C) a monster (FTF) that does not satisfy the antecedent, **satisfies the non-conditional presupposition** and does not satisfy the consequent; and (D) a monster (FF-) that **does not satisfy the antecedent or presupposition** (and as a result the consequent is undefined).

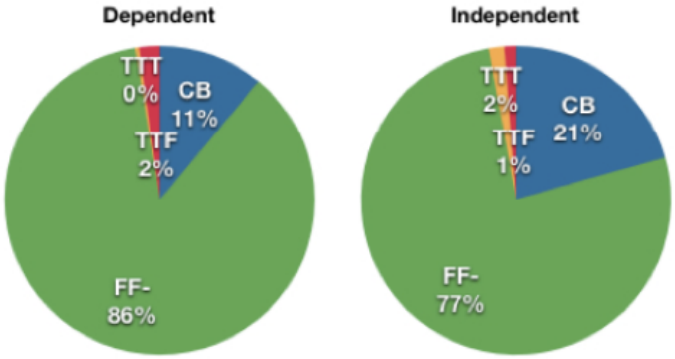
(A) : TTT (B): TTF (C): FTF (D): FF-

(A) : TTT (B): TTF (C): FTF (D): FF-

Romoli et al predicts that in the critical trials, the antecedent *p* of the conditional *if p, then q* is denied by the second sentence *not-p*, which makes the pictures (A) TTT and (B) TTF incompatible with the sentences. Crucially, if the conditional sentence has a non-conditional inference *q*, the picture (C) FTF is the only compatible choice, while if it has a conditional inference $p \rightarrow q$, both (C) FTF and (D) FF- are compatible with the sentences. Under the hypothesis that conditional inferences are more likely when the presupposition of the consequent is dependent on the antecedent, (D) FF- is predicted to be chosen more often in the Dependent condition than in the Independent condition.
As predicted, (D) FF- was chosen significantly more often in the Dependent condition than in the Independent condition. Moreover, in both conditions, (C) FTF was chosen more often than (D) FF-. These results suggest that participants are more likely to make conditional inferences when the presupposition is dependent on the antecedent. The preference for (C) FTF might seem to suggest that the non-conditional inference was preferred in both conditions — this rests on the assumption that participants who arrived at a conditional inference would consistently select FF-. But the conditional inference is in principle compatible with both FF- and FTF. Therefore, while it is certain that a selection of (D) FF- is due to the conditional inference, there is no direct way to know whether (C) FTF is due to a conditional inference or non-conditional.



To address this concern, they also conducted a covered box task (Exp 2), where one of the pictures is covered and cannot be seen. In the critical trials of Exp 2, (C) FTF is covered. Given that the non-conditional inference is only compatible with this picture, while the conditional inference is compatible with both (C) FTF and (D) FF-, they expect that whenever the non-conditional inference arises, the covered picture will be selected; furthermore, whenever the conditional inference arises, (D) FF- will be selected. The results were consistent with those from Exp 1: (D) FF- was chosen significantly more often in the Dependent condition than in the Independent condition. In addition, (D) FF- was selected more often than the covered box in both conditions



The results from Romoli et al suggested that (in)dependence played some role in participants' selection of the inferences that they would endorse for conditional statements like *if p then q*. In particular, in the covered box task, the conditional inference was available in **both** Dependent and Independent conditions: there is a substantial amount of endorsement for (D) FF- in the Independent condition, suggesting that the conditional inference is also available there. This could be due to the nature of covered box task: the general pragmatic consideration that selects the stronger inference (i.e. the non-condional inference *q*) is not observable here, because the selection mechanism is suspended. Thus this is no necessarily incompatible with independence-based accounts which invokes the same consideration to account for selection/preference.

3. Motivation

As we mentioned, our introspection of the classic proviso sentences and the experimental results reported in Romoli et al consistently indicate that (in)dependence does play a role in modulating the availability of the strengthened reading. What has not been answered in this line of work is the following: is it the grammar (following the logic of presupposition projection in conditionals given our theories) or post-semantic reasoning based on world knowledge that underlies the (in)dependence consideration in our assessment of the inferences that arise from the proviso sentences?

Neither our intuition of the classic proviso sentences, nor the results in Romoli et al, can answer this question. To elaborate on the latter, the experimental materials used in

Romoli et al were selected from a norming study, and the items that were included in the Independent condition of their subsequent experiments received a rating of around 2.5 on a scale from 0 to 5 in terms of the relatedness of the antecedent *p* and the presupposition *q*. This property of their materials inherently determined that the (in)dependence manipulation was still muddled by participants' world knowledge, affecting in particular the Independent condition in which the strengthened reading would potentially surface. In addition, this also has important consequences for their experimental results, because the possibility of postulating a dependency relation between *p* and *q* in their design would necessarily lead to an *overestimation* of the conditional inferences, and consequently an *underestimation* of the non-conditional inference, in the Independent condition.

In order to divorce issues of world knowledge and plausibility considerations from the interpretation of conditionals, we propose a study in which key lexical items in the proviso sentences are substituted with pseudowords, which we hope will give us a real shot at understanding what is the semantic presupposition of these sentences.

4. Experiments: The Pseudoword Paradigm & Five Pilot Studies

4.0. The Pseudoword Paradigm

We created a paradigm that uses pseudowords to replace the key lexical items in proviso sentences, and manipulate the (in)dependence relation between *p* and *q*. Within such a paradigm, the critical conditions will take the form of the following:

(1) **Independent**: *if p then qq*, and there is no relation between p and q;
Some people are glorps, and some people are not glorps.
(Glorps or not,) Some people have dords.
If John is a glorp, he will yapple his dord.
Expected inference: John has a dord. (non-conditional)

(2) **Dependent**: *if p then qq*, and an explicit relation between p and q are provided in the context;
Some people are glorps, and some people are not glorps.
{Half of the,Most,All,Only} glorps have dords.
If John is a glorp, he will yapple his dord.
Expected inference: If John is a glorp, he has a dord. (conditional)

Given this paradigm, I ran five pilot studies probing for an independence effect between (1) and (2). The setup of the experiment and the dependent measure used varied across the pilots: to foreshadow the upcoming sections, pilots 1-4 used a story-telling narrative, with the materials involving only one instance of the critical proviso cases, and asked the participants to evaluate how *surprising*/*unexpected* they are to see the negation of the non-conditional inference (e.g. "John doesn't have a dord.") using either a scale from 0-100 or a binary choice; pilot 5 was designed as a logical game involving

four critical trials separated by distractors, and participants were asked to answer a forced choice, polar question about the non-conditional inference (e.g. “Does John have a dord?”). In what follows, I provide more details on each of the pilot studies and report their results.

4.1. Pilot 1 (November 16, 2020)

Design & materials: In Pilot 1, participants were instructed to read four stories about aliens from different planets far away from the Earth, which includes a statement characterizing an individual from each of those planets. For each story, they received two new pieces of information, one at a time. They were asked to evaluate how *surprised* they would be to learn each of these new pieces of information, given what they have been already told about the aliens, by providing a rating for the statements on a scale from 0 to 100, with 0 being not surprised at all and 100 being totally surprised. An example of how they should evaluate *surprised* is provided:

> “Suppose you had previously been told that *Katie recently got divorced*. You might be very surprised if you later found out that *she had never been married*, but not surprised at all to learn that *Katie was unhappy with her marriage*.”

There were a total of four trials, presented in the following order: Control, Independent, HalfDependent, and FullDependent. The Control trial does not include any presupposition triggers, and its two statements involve a contradiction followed by a repetition of what has been mentioned in the context; this trial is used to check if the participants are paying attention to the instruction and/or the task. In the Independent trial, the context explicitly specifies that *p* and *q* are not related (*“Some of the aliens have dords, but it doesn’t matter whether they are glorps or not.”*), whereas in the two Dependent trials, an explicit, statistical correlation between are *p* and *q* provided. In the critical trials, the participants were first asked to evaluate the negation of the non-conditional inference (*“John doesn’t have a dord”*), and then the antecedent of the proviso sentence (*“John is a glorp”*). The latter is used as a secondary attention check, since it is expected to receive the same rating across all critical trials.

	Context	**Statements**
Control	Eva is an alien from this planet. Here, some aliens are fleppers and some aliens are not fleppers. Most of the fleppers have a lammor. **Eva is a flepper, and she has a lammor.**	1. Eva doesn't have a lammor. 2. Eva is a flepper.
Independent	Ann is an alien from this planet. Here, some aliens are glorps and some aliens are not glorps. Some of the aliens have dords, but it doesn’t matter whether they’re glorps or not. **If Ann is a glorp, she will yapple her dord.**	1. Anna doesn’t have a dord. 2. Anna is a glorp.

	Context	Statements
HalfDependent	Bea is an alien from this planet. Here, some aliens are morties and some aliens are not morties. Half of the morties have a lealo. No one else has lealos. **If Bea is a mortie, she will wegget her lealo.**	1. Bea doesn't have a lealo. 2. Bea is a mortie.
FullDependent	Cece is an alien from this planet. Here, some aliens are gogopos and some aliens are not gogopos. All gogopos have a hoopler. No one else has hooplers. **If Cece is a gogopo, she will dippefy her hoopler.**	1. Cece doesn't have a hoopler. 2. Cece is a gogopo.

Logic of the task & Predictions: The prompt question "how surprised would you be to learn that…" probes the participants' commitment to the non-conditional inference in the context. Namely, if they think that *q*, then they will be more surprised to learn later that *not-q*. The more surprised they are to the new statement *not-q*, the more committed they were to q given the context. If we are able to induce the effect of independence on strengthening after removing world knowledge and *a priori* plausibility considerations, then we should expect a difference between Independent and Dependent trials on the rating measures: participants will be more surprised to learn the new statement *not-q* in the Independent trial than the Dependent trials.

More specifically, in the Independent trial, *p* and *q* are unrelated. If the inference has been strengthened to the non-conditional *q*, then participants will add to the context *q*. If so, they should be more surprised to learn later that *not-q*. On the other hand, if our independence manipulation does not lead to strengthening, then participants will only commit to a conditional inference *if p then q*. In that case, we should observe similar ratings of how surprising people are upon learning *not-q* across the different critical conditions. If an independence effect is found in this paradigm, it will suggest that the role of independence on strengthening remains operative even when we remove world knowledge considerations. It is the grammar — following the logic of presupposition projection in conditionals given our theories — that underlies the (in)dependence consideration in our assessment of the inferences that arise from the proviso sentences.

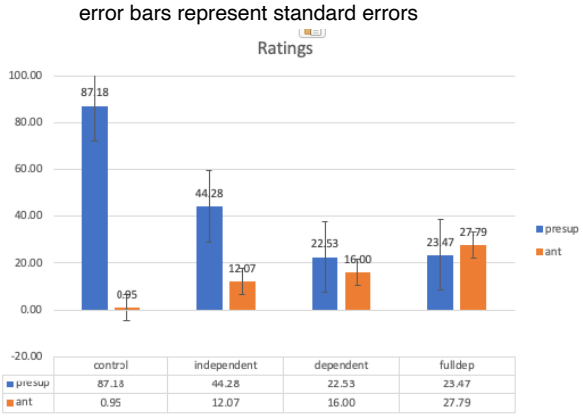
Participants: The link to the experiment was shared on social media, and a total of 22 volunteers participated in the study. Three participants who responded below 60 to the contradiction in the Control trial were excluded. This leaves us with 19 participants.

Results: Ratings for each statement in each trial in plotted in Figure 1. In the Control trial, the contradiction statement received a rating of 87.18, and the repeated statement received a rating close to 0. In each of the critical trials, the blue bars (presup) represent ratings for the negation of the non-conditional inference ("John doesn't have a dord."), and the orange bars (ant) represent ratings for the antecedent of the proviso sentence ("John is a glorp.").

We are primarily interested in the ratings for the negation of the non-conditional inference across the critical trials. In this case, the Independent trial yielded an average

rating of 44.28, and the two Dependent trials yielded an arrange rating of around 23. That is, we see at least a numerical trend toward an independence effect: People were more surprised to learn that "John doesn't have a dord" in Independent trial than they were in the Dependent trials, suggesting that the strengthened, non-conditional inference "John has a dord" was much more available in the Independent trial. By contrast, in the Dependent trials, the ratings for the negation of the non-conditional inference were similar to the ratings for the antecedent of the proviso sentence, indicating that both were at the baseline of how surprised participants would be when they were simply learning or confirming a new piece of information that is consistent with the context.

Figure 1: results from pilot 1;
error bars represent standard errors



Overall, the results in pilot 1 appeared promising. However, the data was collected very informally, and no participant information was collected (e.g. whether their native/first language is English, whether they are linguists, etc). We decided to move on to a more formal method of data collection from naive participants, with a close variant of pilot 1.

4.2. Pilot 2 (November 20, 2020)

Design & materials: A few changes were implemented in this pilot. First, since the task involved pseudowords which may be quite unnatural to naive participants, we included a narrative to make the task slightly more natural, and to justify the use of pseudowords while masking the true purpose of the experiment:

> "Imagine that you are a reporter on extra-terrestrial matters trying to learn things about life outside our galaxy. The aliens on other planets have a different culture and a different language, so you may encounter words that you have never heard of in English -- Don't worry about that, because we are more interested in how you understand the situations."

The proviso sentence (*"If John is a glorp, he will yapple his dord."*) was provided as part of the information that the participant has collected about an alien individual from the planet before an interview. Participants were then asked to evaluate two new statements, the negation of the non-conditional inference (*"John doesn't have a dord"*) and the antecedent of the proviso sentence (*"John is a glorp"*). Each new statement that the participants had to evaluate was presented as a new piece of information that the they *may potentially learn during their interview* with the alien individual.

In addition to the four trials in pilot 1, we added a *Null Context* trial in which no explicit relation of any sort between *p* and *q* was ever mentioned, which appeared immediately after the Control trial. The purpose of this trial is to probe the extent to which the non-

conditional inference would be available when no (in)dependence relation was provided in the context. In retrospect, this condition may have been more appropriate as actually ensuring Independence, if we assume that when no relation between *p* and *q* was provided in the context to the participants, no relation would be assumed on their part, since they (i) could not use world knowledge to postulate relations between pseudowords, and (ii) have no reason to assume a relation between *p* and *q* when none is mentioned in the context.

	Context	Target sentence
NullContext	On this planet, some aliens are tulvorrs and some aliens are not tulvorrs. **You notice the word 'feesp' on a guidebook, but no one explains it to you.** You are about to interview Diane, an alien from this planet. *The information you've collected about Diane so far suggests this:* **If Diane is a tulvorr, she will varkle her feesp.**	1. Diane doesn't have a feesp 2. Diane is a tulvorr.

Finally, we were concerned that the wording of the prompt involving "surprised" may be too vague and perhaps confusing for naive participants, and decided to change it to "unexpected", with the idea being that if something is unexpected, it is inconsistent with participants' belief given what has been said in the context. Apart from these changes, everything else was the same as pilot 1.

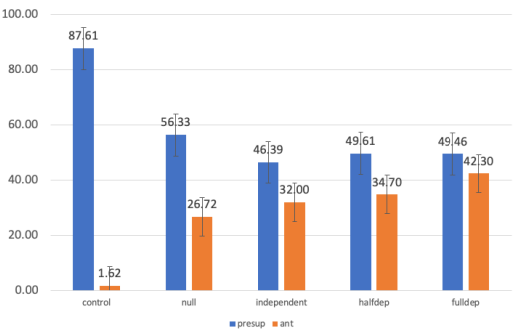
Participants: We recruited 42 participants from MTurk. Exclusion rates were high: three non-native speakers were excluded; in addition, filtering based on the control item (Control Contradiction > 60 AND Control Repetition < 10) excluded another 16 participants. This leaves us with 23 participants.

Results: Ratings for each statement in each trial in plotted in Figure 2. In the Control trial, the contradiction statement received a rating of 87.61, and the repeated statement received a rating close to 0. In each of the critical trials, the blue bars (presup) represent ratings for the negation of the non-conditional inference ("John doesn't have a dord."), and the orange bars (ant) represent ratings for the antecedent of the proviso sentence ("John is a glorp.").

Two major observations can be made upon visual inspection of Figure 2. First, for the negation of the non-conditional inference, while the Null Context trial received a rating of 56.33, which is only slightly higher than the other critical trials, there was no meaningful difference between Independent and Dependent trials.

Second, more seriously, we observed a steady increase for the rating of the antecedent of the proviso sentence, indicating a spillover effect

Figure 2: results from pilot 2;
error bars represent standard errors



across trials. Since the spillover effect might have occurred not only in the evaluation of the antecedent sentence, but also in the evaluation of the negation of the non-conditional inference, this largely made the results from pilot 2 uninterpretable, and calls for a between-subject design.

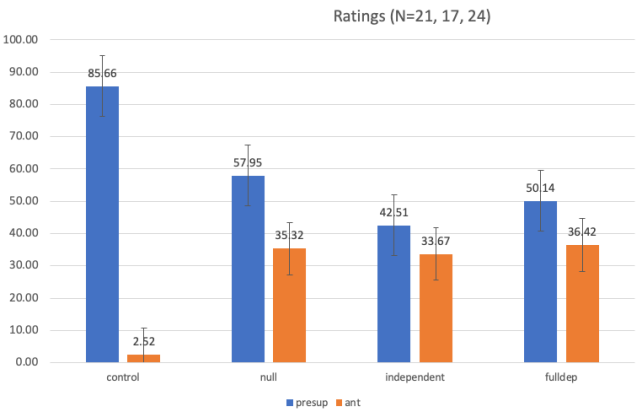
4.3. Pilot 3 (November 24, 2020)

Design & materials: The major change from pilot 2 to pilot 3 is that Independence is now a *between-subject* factor, i.e. each participants will only see the Control trial plus one Critical, which may be either NullContext, Independent, or Dependent. We also changed "unexpected" back to "surprised", based on the feedback that the negative adjective might in fact have made the task more difficult and the intuition less clear compared to our earlier versions. Finally, we also simplified the narrative of the task, such that it was still disguised as a journalistic adventure but less elaborative.

Participants: A total of 119 participants recruited from MTurk for the three subexperiments. Exclusion rates were once again as high as around 50%: three non-native speakers were excluded; a more generous filtering based on the Control trial (Control Contradiction > 50 AND Control Repetition < 20) still excluded 54 more participants, leaving us with only 62 participants (21 in NullContext, 17 in Independent, and 24 in Dependent).

Results: Ratings for each statement in each trial in plotted in Figure 3. In the Control trial, the contradiction statement received a rating of 85.66, and the repeated statement received a rating close to 0. In each of the critical trials, the blue bars (presup) represent ratings for the negation of the non-conditional inference ("John doesn't have a dord."), and the orange bars (ant) represent ratings for the antecedent of the proviso sentence ("John is a glorp.").

Figure 3: results from pilot 3; error bars represent standard errors



Two major observations can be made upon visual inspection of Figure 3. First, for the negation of the non-conditional inference, the Null Context trial received a rating of 57.85, which is again only slightly higher than the other critical trials, and there was a small numerical trend of a difference between Independent and Dependent trials in the opposite direction of what we expected. Second, the rating of the antecedent of the proviso sentence were very similar across trials, suggesting that our concern about the spillover effect in pilot 2 was a reasonable one. The between-subject design allowed us to circumvent the potential spillover effect across trials, and had the advantage of making our results less likely to be influenced by participants strategizing over the course of the experiment. Meanwhile, however, we might also have run the risk of being severely underpowered in

such a between-subject design, and getting participants who have largely different intuitions about how to respond to a presupposition projected out of a conditional with no direct way of comparing how they may respond differently depending on the context.

4.4. Pilot 4 (December 2020)

Design & materials: One pressing issue from the previous two pilots with native participants is the high exclusion rates, which we aimed to address by including a training session at the beginning of the experiment. In pilot 4, we included a training session which providing examples and explicit feedback on how to respond to a statement that is (i) contradictory to the context, which should be surprising, (ii) consistent and contextually implied, which should be not surprising, and (iii) completely new but not inconsistent with the context, which should also be not surprising.

Evaluating the pros and cons of the fully between-subject design, we decided to have a mixed design such that participants either see a Control trial plus the Null Context trial, or a Control trial plus the Independent trial followed by the Dependent trial. We hoped to use this to control for the potential order effects to some extent, especially those between the Null Context trial and the Independent trial, while allowing us to see how the same participant would respond differently to the independence manipulation.

Finally, in this pilot, we decided to collect two types of dependent measures: in addition to collecting ratings on a 0-100 scale, we also created a version in which a forced choice decision had to be made between whether or not the participant was surprised to see a new statement, by responding either Yes or No.

Participants: Including the training session successfully reduced the exclusion rate in each of the subexperiments. Details are summarized in the table below:

	Participants recruited	Participants excluded	Remaining participants
Binary choice with Null Context	28	9	19
Binary choice with Independent + Dependent	25	5	20
Scale with Null Context	24	5	19
Scale with Independent + Dependent	26	6	20

Three non-native speakers were excluded across the four subexperiments; all other participants were excluded based on their performance on the Control trial.

Results for Forced Binary Choice:

Yes-response rates for each statement in each trial in plotted in Figures 4 and 5. In the Control trial, the contradiction statement received 100% Yes responses, and the

repeated statement received 0%. In each of the critical trial, the orange bars (presup) represent the Yes-response rates for the negation of the non-conditional inference (*"John doesn't have a dord."*), and the blue bars (ant) represent the Yes-response rates for the antecedent of the proviso sentence (*"John is a glorp."*).

For the negation of the non-conditional inference, the Null Context received a Yes-response rate of 85.71%, which is particularly high and may indicate that the non-conditional inference is strong. However, the Independent trial only received a 40% Yes-response rate, and the Dependent trial received a 50% Yes-response rate, a numerical trend in the opposite direction of what we would expect. This pattern makes it difficult to evaluate how one should interpret the high Yes-response rate in the Null Context. In retrospect, one possible explanation is the only the Null Context trial ensured "independence" by not making salient any relation between *p* and *q*, whereas even in Independent, providing information such as "Glorps or not, some people have a dord" allows for the possibility of establishing some sort of relation. Furthermore, there is a considerable amount of variability in the Yes-response rates for the antecedent of the proviso sentence; without a clear pattern, it is also unclear how this variability should be interpreted.

Figure 4: results from pilot 4 (binary - NullContext); error bars represent standard errors

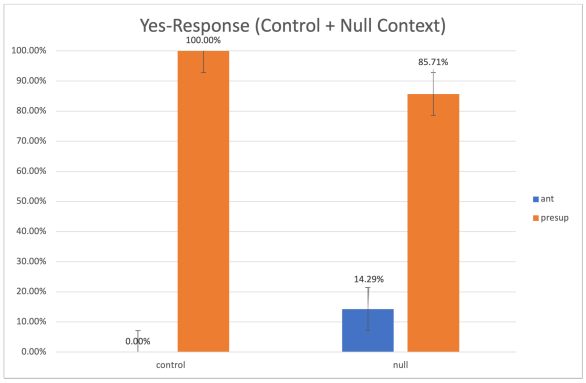
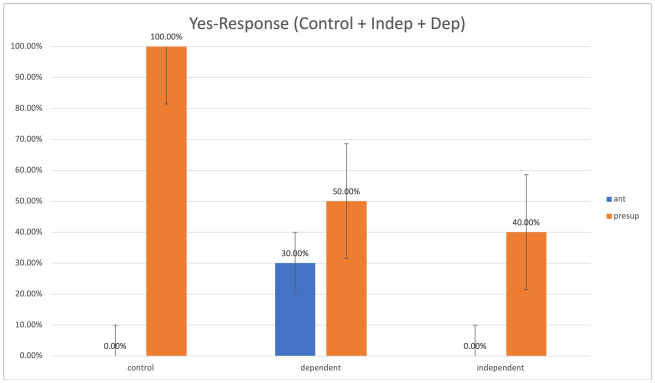


Figure 5: results from pilot 4 (binary (In)Dependent); error bars represent standard errors



Results for Ratings based on 0-100 Scale:

Ratings for each statement in each trial in plotted in Figures 6 and 7. In the Control trial, the contradiction statement received ratings of 85.75 and 88.67, and the repeated statement received ratings of 1.68 and 6.46. In each of the critical trials, the orange bars (presup) represent ratings for the negation of the non-conditional inference ("John doesn't have a dord."), and the blue bars (ant) represent ratings for the antecedent of the proviso sentence ("John is a glorp.").

For the negation of the non-conditional inference, the Null Context trial received a rating of 52.99, the Independent trial 54.31, and the Dependent trial 58.70. There is visually no or little difference between all three critical trials, suggesting a lack of independence effect. Noticeably, there was no difference between the Null Context and the other two trials, contrary to the binary choice task.

Figure 6: results from pilot 4 (binary - NullContext); error bars represent standard errors

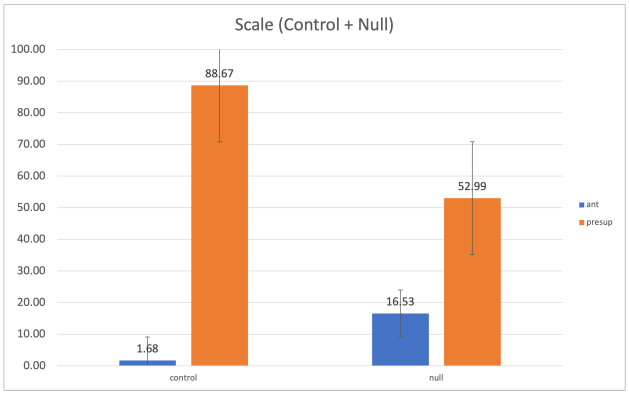
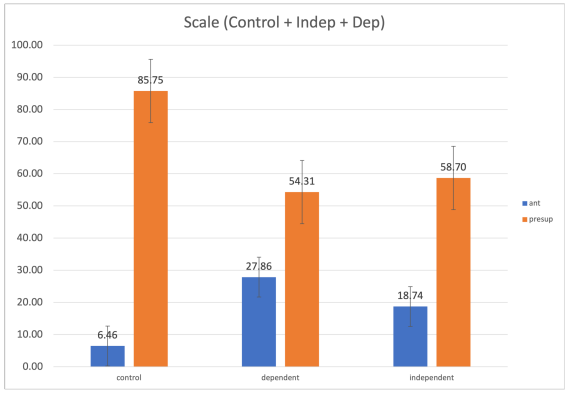


Figure 7: results from pilot 4 (binary (In)Dependent); error bars represent standard errors



One puzzling aspect of the data concerns the convergence and the divergence of the two dependent measures. Typically, we would expect a correlation between the binary choice results and the ratings; a high Yes-response rate should translate into a high rating. For the Independent and Dependent trials, there is a lack of meaningful difference in both measures, with Dependent numerically exceeding Independent by a very small amount on both counts. However, for the Null Context trial, the exceptionally high Yes-response rates in the binary choice task did not actually translate into a high rating on the 0-100 scale. We note, at the same time, that for all the pilots that we conducted with naive participants on MTurk, ratings for all critical trials hovered around 50. This raises a potential concern about the sensitivity of the this dependent measure: given that we are probing a rather nuanced difference in participants' intuition, the scale-based dependent measure may not allow us to detect such a difference, since participants might strategically choose the midpoint of the scale whenever they felt uncertain and decided to move on. Thus, the lack of a correspondingly high rating for Null Context may be due to issues orthogonal to our question at hand.

4.5. Pilot 5 (2021 February 1)

Design & materials: The unsuccessful attempts to find an independence effect in the previous pilot call for a radical rethink of the present paradigm. Crucially, we suspect that in the previous setup, the context and the narrative we provided were so elaborate that they may have obscured the task too much, making it too implicit for us to tap into any differences between the critical trials. We reasoned that since obtaining the expected inferences in the proviso sentences required a level of somewhat complex logical reasoning, we could design the task as a logical game in which we would explicitly tell the participants to perform logical reasoning of various sort.

We additionally speculate that the dependent measure we have been using, i.e. ratings on a scale from 0 to 100, is not a good measure to be used in a logical game, in addition to potentially not being sensitive enough for the reason we discussed in the last section. Thus our only dependent measure in pilot 5 will be a forced binary choice task.

After implementing these two changes, participants see instructions like the following:

> "You will be given 20 puzzles to test your logical reasoning. In each puzzle, you will see 3 statements. You will then be asked to answer a question to the best of your ability, given these statements. Select 'Yes' or 'No' to respond.
>
> "The statements may contain made-up words that you have never heard of in English -- this is on purpose: we use these "pseudowords" so you can focus on the logical aspects of the questions (rather than how plausible or reasonable they may be, given your common knowledge)."

The format of the puzzles is as follows: Three statements were presented on separate lines. Statement 1 always takes the form of *"Some people are glorps and some people are not glorps".* Statement 2 introduces another pseudoword like *"dord"*, but the specific format depends on the trial. Statement 3 is the proviso sentence in critical trials, *"If John is a glorp, he will yeapple his dord."*

In terms of the content of the logical task itself, two further significant changes were implemented. First, we reconsidered the form of the "Independent" condition: previously, this trial provided an explicit denial of any relationship between *being a glorp* and *having a dord*, but it could be the case that the mentioning of *glorps* in "*glorps or not, some people have dords*" has already made salient some relation between *being a glorp* and *having a dord*, or allowed for such a relation to be more likely to be imagined in the context. We thus considered something very simple for this trial: *Some people have dords.* For this reason, a separate Null Context trial is no longer needed: since participants cannot draw inferences/connections from world knowledge about *dords* and *glorps* as we told them from the get-go that these are made-up words, nor are they provided with information about the relation between the two pseudo words in the independent condition, the default assumption is that participants would assume no relation unless indicated otherwise. We thus have only two critical conditions, Independent and Dependent, which will be *between-subject*.

Second, in pilot 5 we increased the number of total trials in the experiment: there are 4 critical trials and 16 distractor trials, with the following structure:

> *Distractor *2 + Critical + Distractor *4 + Critical + Distractor *4 + Critical + Distractor *4 + Critical + Distractor *2*

That is, the experiment begins and ends with two distractor trials, and the 4 critical trials are separated by 4 distractor trials between them.

Given the increased number of trials, we counterbalanced the expected Yes/No responses. Additionally, half of the trials involve a context that includes a conditional statement (4 critical trials + 6 distractor trials), and the other half involve a context that doesn't have any conditional statement (10 distractor trials). An example of each trial type is provided here:

	Statement 1	Statement 2	Statement 3	Question
Distractor_NonConditional	Some people are limpals and some people are not limpals.	No limpals like to toogid.	Ann is a limpal.	Does Ann like to toogid?
Distractor_Conditional	Some people are roopers and some people are not roopers.	If Bea is a rooper, she has a hondart.	Bea is a rooper.	Does Bea have a hondart?
Independent	Some people are glorps and some people are not glorps.	Some people have dords.	If Cece is a glorp, she will yapple her dord.	Does Cece have a dord?
Dependent	Some people are glorps and some people are not glorps.	Only glorps have dords.	If Cece is a glorp, she will yapple her dord.	Does Cece have a dord?

We took extra care to consider the form of the question that we ask the participants. Initially, we considered having a separate prompt followed by a statement, e.g. “Is it possible that John has/doesn’t have a dord?”. However, we were concerned that prompts like this may be too weak and runs the risk of always inviting the participants to be as charitable as possible in their reasoning, which may end up inducing the weaker, conditional inference across the board. Further, a separate prompt may invite different interpretations or biases of the prompt itself. Thus, we instead simply use the polar question *“Does John have a dord?”* The basic idea is that by asking straight up *“Does John have a dord?”*, we will put the participants in a position where we assume they know the answer to this polar question, and thus have to figure out what I’m assuming that they know in order to answer the question.

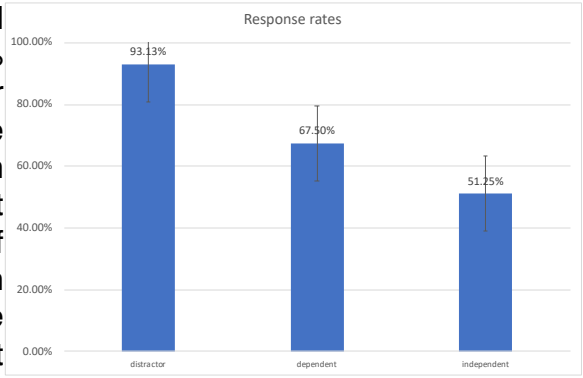
Thus, the best way one can make sense of the question that we are assuming the participants know whether John has a dord or not, given statements 1-3. If the participant can endorse the non-conditional inference *“John has a dord”*, they should respond **Yes**. We further reason that in the case of a conditional inference, *“If John is a glorp, he has a dord”*, participants in principle cannot answer this question with certainty, but they may reason that either (i) it is safer to respond No when the relevant antecedent is unknown, or (ii) since there is no statement confirming John is a glorp, but there always is in the distractor trial, the better option is to assume that perhaps John is not a glop after all, because otherwise the statements would feel under-informative compared to the distractor trials. Following this reasoning, we expect a higher Yes-response rate in the Independent condition, where the non-conditional inference is expected to be available, than the Dependent condition.

Participants: A total of 20 participants were recruited for each subexperiment, giving us 80 observations in each critical condition. Only one participant was excluded for achieving lower than 75% accuracy on the distractor trials.

Results: For the distractor trials, since their expected response involve both Yes and No, we calculated the rates of accurate responses (i.e. accurate if expected Yes and responded Yes, or expected No and responded No; inaccurate otherwise). For the critical trials, Independent and Dependent, we calculated the rates of Yes-responses. These results are plotted in Figure 8.

The distractor trials had 93.13% accurate rates, indicating that the participants were paying attention to the task. Meanwhile, for the critical trials, the Dependent condition yielded a 67.5% Yes-response rate, and is considerably higher than the 51.25% Yes-response rate in the Independent condition. The response rates in the Independent condition is probably at chance; that is, participants responded as if they were guessing. If the response rates in the Dependent condition turns out to be above chance, then this would be opposite of what we would anticipate.

Figure 8: results from pilot 8;
error bars represent standard errors



One possible reason for this reversed pattern could be that there were confounds that we did not control for in the setup of the experiment, which somehow led to the non-conditional inference being even more salient in Dependent than Independent. I will discuss this possibility and what we may learn from it with more details in Section 5.

Another possibility is that contrary to our reasoning of how the participants may respond in the presence of a conditional inference, there were certain other strategies that the participants employed when they were uncertain about how to answer the question. For example, one participant in the Dependent subexperiment left the following comment at the end of the experiment:

> *"I was unsure how to answer the questions that said "IF ____ is a _____" because I was not sure if ____ was a _____ or not, so my answer would have been I don't know, but instead I assumed in each case they were in order to be able to answer the question."*

If such a strategy was prevalent among the participants, then our dependent measures would not be able to detect any effect of independence on the availability of the non-conditional inference, especially if in the Dependent condition the participants simply did not take the contextual information into consideration. However, note also that it cannot be the only explanation for the pattern observed in the data, since the same strategy could be applied in the Independent condition as well, and thus should not lead to any difference between the two critical conditions.

5. Reflections and next steps

The goal of the pseudoword paradigm is to tease apart whether it is the grammar (taken to mean the logic of presupposition projection in conditionals given our semantic theories) or post-semantic reasoning based on world knowledge that underlies the (in)dependence considerations in our assessment of the inferences that arise from the proviso sentences. One hope was that by using this paradigm, we would be able to carefully manipulate and control for what information is (not) available to naive participants, which may give us a clearer and more robust shot at empirically replicating the effect of independence in the proviso sentences, providing the foundation for answering future questions about presupposition projection.

To this end, the pseudoword paradigm was implemented and tested across the five pilot studies reported above. We have not been able to reliably find an independence effect with the tasks and the measures used so far:

- Pilot 1 yielded promising results in the expected direction, but the data was collected very informally and the study was severely underpowered to reach any statistically reliable conclusions;
- Pilot 2 and Pilot 3 yielded null results; in addition, there were issues of interpretability due to the setup of the experiments;
- Pilot 4 yielded, at best, partially interpretable results on one measure and null results on the other measure; considerations regarding the sensitivity of the dependent measure need to be taken more seriously;
- Pilot 5 yielded a potentially meaningful difference but in the opposite direction of what we anticipated.

I do not take the results of these studies to indicate that independence does not play a role in modulating the availability of the (non-)conditional inference; its role is clear from our intuition and from existing data in the literature. Instead, I think there is a lesson we can learn from these failures, a lesson about what factors other than independence that may be at play in regulating the availability of the (non-)conditional inference, not only in our experimental stimuli but more generally in conditionals where we may not expect a strengthened presupposition solely based on independence considerations. Thus, it may very well be the case that we could not find an independence effect because we did not control for other relevant factors that must be in place in order for an independence effect to arise.

I explore one possible explanation for the lack of independence effect in the experiments by reconsidering the Mandelkern examples of what he calls “unexpected strengthening”, and elaborating on some of the responses to these examples. Mandelkern (2016) takes issues with theories that postulate a semantic presupposition for the proviso sentences together with a pragmatic mechanism for strengthening. In these theories, the standard response to the proviso problem is that the semantic presupposition of *“if p then q*$\underline{q}$*”* is the conditional *“if p then* $\underline{q}$*”*, but a speaker is felt to presuppose $\underline{q}$ (i) if *p* and $\underline{q}$ are in some sense independent, and/or (ii) a listener

compares the plausibility of *"if p then q"* and *q*, and concludes that *q* is more relatively plausible. (These conditions can be taken on board together or separately, depending on the theoretical account.)

Mandelkern's objection to these pragmatic explanations of strengthening comes from examples where it appears that the conditional presupposition is strengthened to a non-conditional one even though (i) the presupposition's consequent straightforwardly *depends* on its antecedent, and (ii) there is *no apparent independent source of pragmatic pressure* to strengthen the presupposition. Consider Mandelkern's leading example below, which is supposed to be problematic for pragmatic explanations of strengthening because the speaker is felt to be presupposing *q* rather than *"if p then q"*, and yet none of the usual strengthening mechanisms (i-ii) apply here:

(1) [It is common ground that Smith has gone missing, and we don't know whether he is still alive] A detective enters and says: "If the butler's clothes contain traces of Smith's blood, then it was the butler who killed Smith."
a. If the butler's clothes contain traces of Smith's blood, someone killed Smith
b. ***Someone killed Smith.*** *[felt presupposition]*

Under standard assumptions of satisfaction theories, the semantic presupposition in (1) is (1a). The felt presupposition in (1), however, appears to be the stronger (1b). Mandelkern takes this to indicate that strengthening has occurred, and he contends that this case of strengthening is *unexpected*, because (i) *q* "Someone killed Smith" depends in a straightforward way to the antecedent *p* "the butler's clothes contain traces of Smith's blood", and (ii) it's not clear that there is other pragmatic pressure to opt for *q* over *"if p then q"* based on plausibility considerations (*i.e.* it is perfectly plausible to think that the detective is presupposing *"if p then q"*). As such, the pragmatic accounts of strengthening fail to predict the strengthening felt in (1).

Similarly, in (2-3), the felt presupposition is (2b)/(3b), even though in each case, (i) *q* depends in a straightforward way to the antecedent *p*, and (ii) the context does not renders the conditional (2a)/(3a) implausible or less plausible than (2b)/(3b).

(2) What are the kids up to today?
[I don't know, but] if the kids played baseball, they're the ones who broke the dining room window.
a. If the kids played baseball, the window was broken.
b. ***The window was broken.*** *[felt presupposition]*

(3) Is John in good health? Is he taking care of himself?
[I'm not sure, but we should be able to tell at dinner;] if he's restricting his sugar intake, then his diabetes is under control.
a. If John is restricting his sugar intake, he has diabetes.
b. ***John has diabetes.*** *[felt presupposition]*

Since Mandelkern 2016, there has been some informal/unpublished responses to these examples arguing that they are not in fact counterexamples to the pragmatic solution to strengthening, in the sense that these examples are not dependent so much on "strengthening" of conditional presuppositions or on the presuppositions themselves being non-conditional; they may well be simply orthogonal to the theory of presupposition projection all together. I will provide a sketch of such responses based primarily on my personal communication with Frank Staniszewski and Dan Lassiter.

An important observation that has been made regarding Mandelkern's examples is that all these examples share the same discourse structure: they start with a context with a highly salient, unsettled QUD. This is inferred from (1)'s context but given in (2-3):

(1) Did Smith die? What happen to Smith?
(2) What are the kids up to today? What did the kids do?
(3) Is John in good health? Is he taking care of himself?

However, the conditional presupposition *"if p then q"* in these cases does not directly address these QUDs:

(1) a. If the butler's clothes contain traces of Smith's blood, someone killed Smith.
(2) a. If the kids played baseball, the window was broken.
(3) a. If John is restricting his sugar intake, he has diabetes.

Instead, *"if p then q"* in each of these cases addresses a more specific question that can be posed felicitously only by someone who is assuming that the consequent's presupposition holds. Following this observation, in unpublished work, Staniszewski and Lassiter argue that Mandelkern's examples such as (1-3) are explained by a non-optional process of establishing discourse coherence by inferring why the speaker would have chosen to set up and address a particular QUD (Roberts, 2012). Crucially, utterances invoke QUDs by virtue of their *form*, and the requirement to establish discourse coherence is **non-cancellable**; Mandelkern's generic arguments against pragmatic accounts do not apply in such cases, since they assume that the relevant pragmatic relations are meaning-based and can be overridden by contextual pressures.

Let's assume (following *inter alia* Grice (1989); Starr (2010)) that conditionals have the following discourse-structural characteristics: the antecedent poses, but does not answer, a QUD; the consequent poses and answers a further question within a context subordinated to a *positive* answer to the antecedent QUD. In (2), this amounts to:

(4) If the kids were playing baseball, [*Raise QUD: "Were the kids playing baseball?"*]
they're the ones who broke the dining room window. [*Assuming 'yes', pose & answer another QUD: "Who broke the dining room window?"*]

This approach suggests an account of why the speaker is felt to be presupposing *"the dining room window is broken"* in (2). Consider two scenarios. (i) The speaker is **ignorant** and has no idea whether *p*. In that case, it is very difficult to rationalize their

choice to pose this particular question about that particular window, or to subordinate it to the antecedent QUD. The observed choice of the second QUD is bizarre in this scenario. (ii) The speaker is **informed** and does know that *p*. Then, it is very easy to explain why they would have chosen this particular series of QUDs. Accommodating the information that *p* is thus needed to render the discourse coherent, during this move of QUDs. This results in a salient inference *q*. This reasoning extends to Mandelkern other examples such as (1) and (3), all of which share the relevant discourse-structural features.

By contrast, examples like (5) differ because the same kind of discourse coherence can be established by invoking a generalization linking antecedent and consequent, rather than a specific relevant fact. Attributing to a speaker a general assumption 'Monarchies have kings' is sufficient to rationalize her choice to subordinate *"How well known is Buganda's king?"* to a "yes" answer to *"Is Buganda a monarchy?"*.

(5) If Buganda is a monarchy, its king is little known.
   a. ***If Buganda is a monarchy, Buganda has a king.*** *[felt presupposition]*
   b. Buganda has a king.

As many have noted before, conditional presuppositions appear most readily when such generalizations are available. On this QUD-based account, the availability of such generalizations determines whether one can rationalize the subordination move without attributing further assumptions to the speaker.

Furthermore, Dan Lassiter suggests (p.c.) that this sort of reasoning may also apply to certain well-known examples of non-conditional presuppositions projecting from the consequents of conditionals, subject to independence manipulations. Consider, for example, the classic case where *p* and *q* are dependent:

(6) If Theo is a scuba diver, he'll bring his wetsuit on vacation.

The antecedent in (7) addresses the QUD *"Is Theo a diver?"* We then make a discourse move to addresses a further QUD, *"On the assumption that he is [a diver], will he bring his wetsuit?"* It is not difficult to rationalize this choice of discourse strategy without recourse to the assumption that Theo owns a wetsuit, by appealing to a generalization of the form *Divers own wetsuits*. Of course, it could also be that the speaker knows that Theo owns a wetsuit, but does not know whether he uses it for diving or surfing; if there is no surf in the vacation destination this kind of reasoning would do just as well. In sum, the speaker may well be presupposing that Theo owns a wetsuit, but out of context there is no overwhelming pressure to attribute this belief to him.

In (7), *p* and *q* are supposedly independent, and the felt presupposition is the non-conditional *"Theo has a wetsuit"*:

(7) If Theo is smart, he'll bring his wetsuit on vacation.

Here, the antecedent addresses the QUD *"Is Theo smart?"* Then we make a discourse move to addresses another QUD, *"On the assumption that he is [smart], will he bring his wetsuit?"* It is quite difficult to rationalize this choice of discourse strategy (i.e. moving to the second QUD from the first one) without recourse to the assumption that Theo owns a wetsuit. The best option that comes to mind is to appeal to a generalization of the form *Smart people own wetsuits*, but since this generalization seems quite weird, we are inclined instead to favor the rationalization that involves attributing to the speaker the specific knowledge that Theo owns a wetsuit.

Let's now circle back to our experimental materials. Setting aside pilot 1 which showed hints of an independence effects and pilot 2 which has issues of interpretability due to the design, I take the results across the 3 latest studies to indicate that in the Dependent condition, somehow the non-conditional inference is more readily available than expected, or that the conditional inference is not as available/detectable.

(8) Some people are glorps and some people are not glorps.
All/Only glorps have dords.
If John is a glorp, he will yapple his dord. [Dependent]

(9) Some people are glorps and some people are not glorps.
Some people have dords.
If John is a glorp, he will yapple his dord. [Independent]

Across these studies, the prompts we have used — though the specific format varies — all made salient an unsettled QUD, "Does John have a dord?" This is precisely the question we used in pilot 5, and is directly related to the evaluation of how surprising it is to learn that "John doesn't have a dord" in previous pilots. Importantly, this is reminiscent of the discourse structure in the Mandelkern examples in (1-3), where the context made salient a QUD, which cannot be settled by the conditional inference.

More specifically, if we try to apply the reasoning of discourse strategy as sketched above to (8), the antecedent first raises the QUD *"Is John a glorp?"* We then make a discourse move to a further QUD, *"On the assumption that he is [a glorp], will he yapple his dord?"* It is in principle not very difficult to rationalize this choice of discourse strategy without recourse to the assumption that *John owns a dord*, by appealing to a generalization of the form *Glorps own dords.* This generalization is readily available in the Dependent condition where the second statement is "All glorps have dords" (pilots 3-4); perhaps a little less readily available in pilot 5 where the statement took the form "Only glorps have dords". Regardless, this generalization should not be weird or implausible in the given context, and thus there is no need to favor the rationalization that involves attributing to the speaker the specific knowledge that John owns a dord, in order to satisfy discourse coherence. We should expect the conditional presupposition, "If John is a glorp, he has a dord", to be available.

However, this does not provide an answer to the immediate, salient QUD that our prompts pose, "Does John have a dord?" That is, although the conditional

presupposition “If John is a glorp, he has a dord” itself is expected from dependence, it cannot be used to settle the immediate QUD that our prompts set up; the QUD could only be addressed if you assume the antecedent of the conditional is true. And since the requirement to address the immediate QUD and establish discourse coherence is non-cancellable, it can override the independence effect. Note, importantly, that the more classic proviso examples like (6) and (7) do not have the same discourse structure like our experimental materials or the Mandelkern examples; there is no pressing QUD regarding *p* or *q* in either (6) or (7) such that the conditional inference would unavoidably fail to address. Consequently, by using the prompts we have, we might have just accidentally been reproducing the Mandelkern examples in our studies. If so, in retrospect, the participant’s strategy to “assume in each case the antecedent were in order to be able to answer the question” was not coincidental; it reflects the non-cancelled pressure to address the immediate QUD, which the conditional inference cannot settle.

Thus, one way to think about the results from these pilot studies is that independence is not the only relevant factor regulating the availability of (non)-conditional inferences, and its effect can be overwritten by discourse pressure from QUD. By itself, independence is not sufficient as a solution to determining which inference to select in the proviso problem. Our prompt question may have served as the salient QUD that needs to be resolved, and in the case that dependence only allows for a conditional inference, the QUD could only be resolved by assuming the antecedent and the consequent are both true. (That is, if we assume the antecedent is false, we still cannot settle the QUD.)

Following this, regardless of whether or not we stick to the pseudoword paradigm (and perhaps we shouldn’t), one thing that needs to be taken more seriously into account in future work is the role that a salient QUD in the discourse may play, which likely will interact with independence factors. In a sense, though, what we have considered as (in)dependence effects can be recast a byproduct of the requirement of evaluating the kind of assumption that needs to be added in order to rationalize a choice of moving from one QUD to the next: a general assumption, which follows from some well-known or plausible dependence relations about *p* and *q* based on our world knowledge, is sufficient for rationalizing the discourse strategy that ensures coherence, allowing for the conditional inference to survive since the stronger non-conditional inference is not required; in the absence of such a general assumption — a scenario that coincides with independence — the listener favor attributing to the speaker a very specific assumption, *q.*, in order to rationalize the discourse move. Theorizing the kind of role that QUD plays in modulating the availability of the (non-)conditional inferences in the proviso sentences will be an important next step, in order to disentangle issues of projection *vs.* accommodation (strengthening).